Project Proposal : Machine Learning (CAP6610)

Vikas Chaubey
UFID : 3511 5826
Email : vikas.chaubey@ufl.edu
Department of Computer and Information Science and Engineering
University of Florida
Gainesville, Fl 32608

1 Team Introduction

I have decided to do this project independently in order to learn and understand the end to end process of development of neural networks in actual practice.

2 List of Research Papers for Project Reference

- **Learning Deep Features for Discriminative Localization :** By Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, Antonio Torralba Computer Science and Artificial Intelligence Laboratory, MIT
- **Self-taught object localization with deep networks :** By A. Bergamo, L. Bazzani, D. Anguelov, and L. Torresani.
- **Integrated recognition, localization and detection using convolutional networks :**By P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun.
- **Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks :** By Aditya Chattopadhay ; Anirban Sarkar ; Prantik Howlader ; Vineeth N Balasubramanian

3 Introduction and Background

The project aims at implementing a convolution neural network which can perform task of object detection within a given digital image.Object detection techniques are used to identify objects in a given image , classify them and locate them with a bounding box within image.The project focuses on implementing a neural network using weakly supervised object detection techniques.

What is Object Detection?
object detection is a task in computer vision that involves identifying the presence, location, and type of one or more objects in a given image/video.It is widely used in computer vision tasks such as image annotation, activity recognition, face detection, face recognition, video object co-segmentatiom etc.When humans look at images or video, we can recognize and locate objects of interest within a matter of moments. The goal of object detection is to replicate this intelligence using a computer program.

An Object detection neural network has to perform following tasks:

1. object recognition (recoginizing the presence of objects in a given image/video)
2. Object classification (classifying the category of different objects)
3. Localization (determining the location of the objects in the image by putting a bounding box around them)

Methods for object detection generally fall into either machine learning-based approaches or deep learning-based approaches. For Machine Learning approaches, it becomes necessary to first define features using one of the methods below, then using a technique such as support vector machine (SVM) to do the classification. On the other hand, deep learning techniques are able to do end-to-end object detection without specifically defining features, and are typically based on convolutional neural networks (CNN).

Machine Learning approaches:

1. Viola–Jones object detection framework based on Haar features
2. Scale-invariant feature transform (SIFT)
3. Histogram of oriented gradients (HOG) features

Deep Learning approaches:

1. Region Proposals (R-CNN,Fast R-CNN, Faster R-CNN)
2. Single Shot MultiBox Detector (SSD)
3. You Only Look Once (YOLO)
4. Retina-Net
5. Deformable convolutional networks

In recent years, deep learning techniques are achieving state-of-the-art results for object detection, such as on standard benchmark datasets and in computer vision competitions. Notable is the "You Only Look Once," or YOLO, family of convolutional Neural Networks that achieve near state-of-the-art results with a single end-to-end model that can perform object detection in real-time.

What is weakly supervised learning? Most of the deep learning approaches for object detection require that the objects should be annotated within bounding boxes in the training data sets in order to train model.This is an effective approach to train models with with high accuracy.But in case of very large data sets cost of labelling bounding boxes annotations around objects is very expensive.also annotations produced by humans are subjective.Hence weakly supervised object detection techniques could be used to train models which do not require training data to be bound boxed.In these techniques only labelled data without bounding box annotation could be used for training of models which can reduce data preparation effort and reduce development costs.

the main reason and motivation behind choosing object detection for my project is that computer vision as a field has always fascinated me.The wide range of use cases in different fields makes this domain really interesting for example object detection is currently being used to build self driving cars (built by Tesla), Delivery drones (being developed by Amazon) , face recognition and Pedestrian tracking etc.I think this project will give me good exposure to learn and understand the advancements in the field of object detection.

4 Importance of Research Papers

The above mentioned research papers describe techniques with which weakly supervised learning could be implemented without using annotated training data.Grad cam algorithm is a gradient based object identification and classification algorithm.These papers describe generation of class activation maps (CAM) using global average pooling(GAP) in CNNs. These class activation maps could be used to identify and classify objects in CNNs without training them with annotated data set.

5 Algorithms and dataset

In order to implement weakly supervised object detection following algorithms are very crucial:

1. GRAD CAM
2. GRAD CAM ++

I plan to use CIFAR-10 data set for training and testing of this model. This dataset is for image classification. It consists of 60,000 images of 10 classes (each class is represented as a row in the above image). In total, there are 50,000 training images and 10,000 test images. The dataset is divided into 6 parts – 5 training batches and 1 test batch. Each batch has 10,000 images.